

Side effect profile and comparative tolerability of 21 antidepressants in the acute treatment of major depression in adults: protocol for a network meta-analysis

Anneka Tomlinson,^{1,2} Orestis Efthimiou,³ Katharine Boaden,² Emma New,² Sarah Mather,² Georgia Salanti,³ Hissei Imai,⁴ Yusuke Ogawa,⁵ Aran Tajika,⁶ Sanae Kishimoto,⁴ Sino Kikuchi,⁴ Astrid Chevance,^{7,8} Toshi A Furukawa,⁴ Andrea Cipriani^{1,2}

1. Department of Psychiatry, University of Oxford, Oxford, UK
2. Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK
3. Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland
4. Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan
5. Department of Healthcare Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan
6. Department of Psychiatry, Kyoto University Hospital, Kyoto, Japan
7. METHODS Team, Center for Research in Epidemiology and Statistics Sorbonne Paris Cité, Paris, France
8. Paris Descartes University, Paris, France

Correspondence to:

Anneka Tomlinson
Department of Psychiatry
Warneford Hospital
University of Oxford
OX3 7JX, Oxford
United Kingdom
Email: anneka.tomlinson@psych.ox.ac.uk

Abstract

Introduction

We have recently compared all second-generation as well as selected first-generation antidepressants in terms of efficacy and acceptability in the acute treatment of major depression. Here we present a protocol for a network meta-analysis aimed at extending these results, updating the evidence base and comparing all second-generation as well as selected first-generation antidepressants in terms of specific adverse events and tolerability in the acute treatment of major depression in adults.

Methods and analysis

We will include all double-blind randomised controlled trials comparing one active drug with another or with placebo in the acute treatment major depression in adults. We will compare the following active agents: agomelatine, amitriptyline, bupropion, citalopram, clomipramine, desvenlafaxine, duloxetine, escitalopram, fluoxetine, fluvoxamine, levomilnacipran, milnacipran, mirtazapine, nefazodone, paroxetine, reboxetine, sertraline, trazodone, venlafaxine, vilazodone and vortioxetine. The main outcomes will include the total number of patients experiencing specific adverse events; experiencing serious adverse events; and experiencing at least one adverse event. Published and unpublished studies will be retrieved through relevant database searches, trial registries and websites; reference selection and data extraction will be completed by at least two independent reviewers. For each outcome we will undertake a network meta-analysis to synthesise all evidence. We will use local and global methods to evaluate consistency. We will perform all analyses in R. We will assess the quality of evidence contributing to network estimates with the CiNeMA (Confidence in Network Meta-Analysis) web application.

Ethics and dissemination: This review does not require ethical approval.

PROSPERO registration number: CRD42019128141.

Background

Depression affects 350 million people worldwide and it is the second leading cause of global disease burden.¹ The high direct and indirect costs for major depression are substantially due to significant deficits in treatment provision. There are a number of efficacious pharmacological and non-pharmacological interventions for depression, however a significant proportion of patients with major depression remain inadequately treated. Antidepressants are widely prescribed across the world in both primary and secondary care; however, poor adherence and premature discontinuation of antidepressant medication contribute to suboptimal clinical outcomes. Up to one third of patients discontinue antidepressants due to adverse effects and this is a major barrier to antidepressant treatment.

Our recent GRISelda project (Group of Researchers Investigating Specific Efficacy of individual Drugs for Acute depression) reported that the acceptability of antidepressants and drop-outs due to adverse events vary between drugs and the withdrawal rates tend to be higher than placebo.² This current network meta-analysis (NMA) is the completion of the GRISelda project and is based on the same protocol (that have the same PROSPERO registration number, CRD42019128141).³ We have designed this NMA to investigate the profile of specific adverse events for each antidepressants. This will contribute to a better understanding of how to use antidepressants in the treatment of depression in adults.⁴

The objective of this NMA is to compare the specific side effects and the overall tolerability of all second-generation antidepressants and selected first-generation antidepressants in the acute treatment of major depressive disorder in adults. The project is called Meta-Analysis of Relative Tolerability and Harms of Antidepressants (MARTHA).

Methods and analysis

Types of studies

We will include double-blind RCTs comparing one active drug with another or with placebo, as monotherapy, in the acute phase treatment of major depression. Cross-over and cluster randomised trials will be included, while quasi-randomised trials will

be excluded. For cross-over studies, to address concerns around possible 'carry over' effects, we will use data from the pre-crossover phase.³

Types of participants

Patients aged 18 years or older, of both sexes, with a primary diagnosis of unipolar major depression according to standard operationalised diagnostic criteria, such as Feighner criteria, Research Diagnostic Criteria, DSM-III, DSM-III-R, DSM-IV, DSM-5, ICD-10 and ICD-11, will be included. Studies in which 20% or more of the participants may be suffering from bipolar or psychotic depression will be excluded. A concurrent secondary diagnosis of another psychiatric disorder will not be considered as exclusion criterion, but RCTs in which all participants have a concurrent primary diagnosis of another mental disorder or concomitant medical disorder will be excluded. Antidepressant trials in depressive patients with a serious concomitant medical illness, post-partum or treatment resistant depression will be excluded.

Types of interventions

We will include the following antidepressants: agomelatine, amitriptyline, bupropion, citalopram, clomipramine, desvenlafaxine, duloxetine, escitalopram, fluoxetine, fluvoxamine, levomilnacipran, milnacipran, mirtazapine, nefazodone, paroxetine, reboxetine, sertraline, trazodone, venlafaxine, vilazodone and vortioxetine (see GRISELDA protocol for more detail).³ Rescue medications will be allowed if equally provided among the randomised arms. We will include only studies randomising patients to drugs within their licensed dose range.^{3,5} We anticipate that any patient who meets all inclusion criteria could, in principle, be randomised to receive any of the interventions in the synthesis comparator set (assumption of transitivity).

Outcome measures and categorisation of adverse events

Tolerability will be evaluated using the following outcome measures:

1. Total number of patients experiencing one specific adverse event;
2. Total number of patients experiencing serious adverse events;
3. Total number of patients experiencing at least one adverse event.

Two independent researchers will extract all adverse effects reported in the trials (paying careful attention not to double count events) and will then use preferred terms from MedDRA (<https://www.meddra.org/>) to categorise each adverse event. MedDRA has been developed by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use to provide a single standardised international medical terminology which can be used for regulatory communication and evaluation of data pertaining to medicinal products for human use. As a result, MedDRA is designed for use in the registration, documentation and safety monitoring of medicinal products through all phases of the development cycle (i.e., from clinical trials to post-marketing surveillance).

There are five levels to the MedDRA hierarchy, arranged from very specific to very general. At the most specific level, called “Lowest Level Terms” (LLTs), there are more than 70,000 terms which parallel how information is communicated. These LLTs reflect how an observation might be reported in practice (i.e. in a specific study). This level directly supports assigning MedDRA terms within a user database. Each member of the next level, “Preferred Terms” (PTs), is a distinct descriptor (single medical concept) for a symptom, sign or disease diagnosis. Each LLT is linked to only one PT and each PT has at least one LLT (itself) as well as synonyms and lexical variants (e.g., abbreviations, different word order). If we find different MedDRA terms to identify similar adverse events, these synonyms will be merged using clinical judgement into broader categories (as applicable) and validated by another clinician. Any discrepancies will be solved by consensus within the review team.

To define serious adverse events, we will use the classification employed by the US Food and Drug Administration (<https://www.fda.gov/>):

- results in death, or
- is life-threatening, or
- requires inpatient hospitalisation or causes prolongation of existing hospitalisation, or
- results in persistent or significant disability/incapacity, or
- may have caused a congenital anomaly/birth defect, or
- requires intervention to prevent permanent impairment or damage.

All serious adverse effects will be included in the meta-analysis.

Common and very common adverse events

We will also identify common and very common non-serious adverse events using the approved definition of frequency of adverse event issued by the Council of International Organizations of Medical Sciences:

Type of adverse event	Frequency
<i>Very common</i>	$\geq 10\%$
<i>Common</i>	$\geq 1\%$ and $< 10\%$
<i>Uncommon</i>	$\geq 0.1\%$ and $< 1\%$
<i>Rare or very rare</i>	$< 0.1\%$

If the number of common and very common adverse effect is over 20, a survey including patients and clinicians' perspective will be carried out to select which adverse effects should be considered for use in the in the statistical analyses.

Survey

We will include two types of participants:

- Patients: any individual over 18 years old with a current/previous episode of unipolar depression and current/previous use of antidepressants.
- Prescribing clinicians: any healthcare professional (psychiatrist, general practitioner, prescribing nurse or prescribing pharmacist) with personal experience in prescribing and monitoring antidepressants in depression.

The questionnaire will be in English, French and German. We aim to recruit at least 200 patients and 100 physicians from multiple countries to increase the external validity of the findings.

The survey will collect data about the following aspects:

1) *Sociodemographic characteristics and health status of participants:*

For patients: sex, age, country of residency, number of years of education, income, PHQ-9, diagnosis, setting of care, number and names of antidepressants ever taken, duration of treatment, suicidal behaviour, length of current episode, total length of exposure to antidepressants.

For prescribing clinicians: sex, age, country of practice, profession, duration of clinical experience, workplace, personal experience of depression/antidepressants.

2) Ranking of the adverse events

Each participant will be asked to rank the adverse event according to their personal preference. The list presented to patients will contain only clinical adverse events that can be understood by lay people, whereas the list of adverse events for clinicians will also include biological measures (for instance, liver function or glucose blood levels). For patients, we will use the specific “patient-friendly” wording of MedDRA and for clinicians the MedDRA terminology.

A modified Q-sort method will be used to rank the adverse events and the final list of adverse events will include all of the serious adverse events plus the 20 most important non-serious adverse events. If appropriate, the researchers will review the lists of adverse events generated from the survey and include any further adverse events considered to be clinically relevant.

Search strategy and study selection

We will use the same search strategy that we used before for GRISELDA³ and perform an update of the search. The reference selection process will be done by two researchers independently. Any disagreements will be resolved via discussion with a third member of the review team.

Data extraction

Two reviewers will independently extract from the included studies the relevant information about specific adverse events using a predefined structured template. Any discrepancies will be discussed between the two reviewers and any unresolved discrepancies will be resolved by a third senior reviewer. When different values are provided in the published and unpublished studies, the unpublished data will be

prioritised and extracted. Two review authors will ascertain that the data are entered correctly into the final data set.

Length of trial

We will consider the number of participants with adverse events in each treatment arm at 8 weeks.⁵ If information at 8 weeks is not available, we will use data ranging between 4 and 12 weeks (we will give preference to the timepoint closest to 8 weeks; if equidistant, we will take the longer outcome). Longer term studies will be included in the systematic review but excluded from the statistical synthesis of data if they do not provide data for the 4–12 weeks period.

Comparability of dosages

We will include only study arms randomising patients to drugs within the licensed dose. Both fixed-dose and flexible-dose designs will be allowed.

Risk of bias assessment

We will assess risk of bias in the included studies using the tool described in the Cochrane Collaboration Handbook as a reference guide (<http://handbook-5-1.cochrane.org/>). The assessment will be performed by two independent raters. If the raters disagree, the final rating will be made by consensus with the involvement (if necessary) of another member of the review group.

Statistical synthesis of study data

We will generate descriptive statistics for the trial, and study population characteristics across all eligible trials, describing the types of comparisons and some important variables, either clinical or methodological (such as year of publication, age, severity of illness, sponsorship and clinical setting). We will draw the network diagram to graphically present the available evidence.

Pairwise meta-analyses

For each pairwise comparison in the dataset that is informed by 10 studies or more, we will synthesize data using a random effects meta-analysis model, to obtain ORs and 95% Confidence Intervals (CIs). This model assumes that true underlying

treatment effects are similar, but not identical across the different study settings, and allows us to estimate heterogeneity.

One complication we expect to face is that for many specific adverse events we may have low or very low event rates in our dataset. When the outcome is rare, i.e. when there are studies with zero arms in one or both treatment arms, the inverse-variance method for meta-analysis might lead to biased results (<https://training.cochrane.org/handbook/version-6/chapter-10-draft>). In such cases we will also use the Mantel-Haenszel method to synthesize the evidence.⁶ This model avoids the use of the so-called 'continuity correction', which artificially imputes data and might bias results. The model assumes a common (fixed) treatment effect, i.e. does not include heterogeneity. This is a limitation of the approach, but, as the Cochrane Handbook suggests, incorporation of heterogeneity should be a secondary consideration when attempting to estimate treatment effects from sparse data (<https://training.cochrane.org/handbook/version-6/chapter-10-draft>). In order to decide which method to use as our primary analysis for rare outcomes, we will fit a fixed-effects inverse-variance and a Mantel-Haenszel model, and compare results. If the two approaches provide similar results, we will conclude that the continuity corrections have minimal effect on the results of the inverse-variance method. In that case, we will employ a random-effects inverse-variance model as our primary analysis. If there are important discrepancies between the two approaches, we will only use the Mantel-Haenszel method.

Furthermore, when data are rare, the choice of model becomes important,⁷ and different models might give substantially different results. Thus, for the five most important rare outcomes according to our ranking, we will employ additional models (Peto odds ratio, a Bayesian meta-analysis model with informative prior distributions for heterogeneity⁸ and a beta-binomial model, as seen fit according to the assumptions of the different models).⁷ This will allow us to assess the robustness of our findings under different model choices. If different models lead to substantially different results, we will present all results on equal grounds and we will refrain from drawing firm conclusions regarding relative treatment effects.

For all pairwise meta-analyses we will present forest plots. We will use a 0.5 continuity correction, in order to present in the plots studies with zero events in one of their arms. For studies with zero events in both arms we will not show any relative effects.

We will visually inspect the forest plots to identify any particularly heterogeneous comparisons. For the analyses where random effects model will be used, we will compare the estimated standard deviation of random effects with the corresponding empirical distribution.⁸ We will also report the I^2 statistic and its 95% CI, as an additional measure of heterogeneity in the pairwise meta-analyses.

Assessment of the transitivity assumption of NMA

The key underlying assumption of NMA is the assumption of transitivity.^{9,10} In order to assess the validity of this assumption, we will investigate whether study-level characteristics that may impact on the relative treatment effects (i.e. effect modifiers) are similarly distributed across treatment comparisons. Potential effect modifiers include clinical and demographic characteristics, such as age, gender, dose and severity of symptoms. We will group studies by treatment comparisons and obtain descriptive statistics regarding these important covariates. In case we find significant discrepancies in the corresponding distributions, we will limit our network meta-analyses to studies that are sufficiently similar.

The clinical features, which have been demonstrated to date to moderate efficacy of antidepressants include bipolarity,¹¹ psychotic features,¹² and subthreshold depression.¹³ We have assured transitivity in our network with regard to these variables by limiting our samples to participants with non-psychotic unipolar major depression. Other clinical or methodological variables that may influence our primary outcomes of antidepressant efficacy or acceptability include: age,¹⁴ depressive severity at baseline¹⁵ and the dosing schedule.¹⁶ We will investigate if these variables are similarly distributed across studies grouped by comparison.

Network meta-analyses

If we find no evidence against the transitivity assumption, we will synthesize the evidence using NMA.¹⁰ For non-rare outcomes we will use a random-effects NMA model¹⁷ fit in a frequentist setting, assuming a common heterogeneity parameter across all treatment comparisons. We will present the 'league-table' of results, i.e. a table with all estimated treatment effects and the corresponding 95% CIs. For each outcome, in order to assess the extent of heterogeneity, we will compare the estimated value for the heterogeneity standard deviation with the corresponding empirical distributions.⁸ In addition, we will present the prediction intervals for each drug vs. placebo; this will allow us to gauge the effect of heterogeneity in the true underlying treatment effects of a future study. We will rank the various treatments for each outcome using the surface under the cumulative ranking curve (SUCRA).¹⁸

For rare outcomes, i.e. when there are studies with zero events in some of their treatment arms, we will perform a NMA using a fixed-effects Mantel-Haenszel NMA approach¹⁹ and compare results with the fixed-effects inverse-variance NMA model. If results agree, we will use the random-effects NMA model as our primary analysis.¹⁷ If we find important discrepancies we will only present results from the Mantel-Haenszel NMA approach. In sensitivity analyses we will also employ a NMA model with a non-central hypergeometric (NCH) likelihood.²⁰ Both Mantel-Haenszel and NCH NMA can handle studies with zero events in one (but not all) of their treatment arms; in simulations we have shown that these two models perform well under sparse data settings. Notably, both models exclude studies with zero events in all treatment arms. Thus, for very rare outcomes, we expect that the network might become disconnected. In that case we will perform NMAs in each of the corresponding sub-networks that include enough data to be meaningfully synthesized.

For some specific outcomes (i.e. gastrointestinal side effects, neurological symptoms, etc) we assume that the relative treatment effects of the various drugs vs. placebo are exchangeable, i.e. they follow an underlying common distribution. This is based on the assumption that the different drugs might have similar pathways to the outcome. For these outcomes, we will employ a Bayesian, multi-level hierarchical NMA model that assumes exchangeability of the treatment effects against a common comparator. This model has been shown both theoretically and in simulations to lead to an increase of

the statistical power to detect treatment effects of drugs vs. placebo, while automatically controlling for the possibility of multiple testing issues.²¹

Assessment of inconsistency

Inconsistency corresponds to the (statistical) disagreement between the different sources of evidence in a network.¹⁰ Assessment of inconsistency is an important part of NMA. It offers an additional, quantitative method of exploring the validity of transitivity assumption.⁹ Large inconsistency implies a breach of transitivity, which in turn suggests that synthesizing data in a NMA should be avoided.

We will use two different methods for assessing inconsistency in the network. The first one is a 'global' method, the design-by-treatment test.²² This is a test against the null hypothesis of overall consistency in the network. Subsequently, we will employ a 'local' method, 'Separate Indirect from Direct Design Evidence' (SIDDE).¹⁹ Using SIDDE, we group the studies by design (i.e. according to the group of treatments they compare). Then, for each treatment comparison in each design we estimate the direct evidence (from studies of this particular design) and indirect evidence (from the rest of the network). We then compare the two estimates; important differences will point to hotspots of inconsistency in the network.

If these methods suggest the presence of important inconsistency in the network, we will first try to scan our data for extraction errors. If none is found, we will revisit the studies to assess again the plausibility of the transitivity assumption, especially if some hotspots of inconsistency are identified using the SIDDE approach. If we identify possible reasons for this inconsistency we will account for it by performing subgroup analyses. If we cannot identify the cause of inconsistency, we will refrain from performing a NMA.

All methods for inconsistency, however, are expected to have low power in detecting breaches of the transitivity assumption. Especially for the case of rare outcomes (which we expect to have in our analyses), all tests for inconsistency are expected to be extremely low powered. In addition, absence of a statistically significant result in tests for inconsistency does not offer proof of transitivity. Thus, we aim to perform a

thorough assessment of transitivity even in the absence of any proof of inconsistency.

Exploring heterogeneity and inconsistency and sensitivity analyses

We expect small amounts of heterogeneity and inconsistency to be present given the variety of study settings we plan to include. For the most common adverse events, we will explore whether treatment effects are robust in subgroup analyses and network meta-regression using the following characteristics: (1) study year; (2) sponsorship; (3) depressive severity at baseline; (4) dosing schedule; (5) head-to-head vs placebo controlled studies; (6) single-centre vs multi-centre studies).¹ The sensitivity of our conclusions will be evaluated by analysing (1) only studies with balanced doses in all arms (i.e., we will exclude studies with unfair dose comparisons); (2) only studies with unpublished data (i.e., we will exclude studies providing published data only); (3) only studies with low risk of bias; (4) only head-to-head studies.

Assessing small study effects, publication bias, reporting bias

It has been empirically shown that safety outcomes are in high risk of reporting bias,²³ and that trials tend to systematically understate adverse events.²⁴ This phenomenon might be more pronounced in placebo-controlled trials.² In order to assess the existence of small-study effects and publication biases, we will use funnel plots and contour-enhanced funnel-plots.²⁵ This will allow us to check whether the precision of the studies (which is directly related to sample size) correlates with the effect size. We will use the Harbord test²⁶ to formally test for asymmetries in the funnel plots. We will follow this procedure for pairwise comparisons between antidepressant and placebo. If we identify an important association of the reported effect with the trials' precision, we will try to adjust for it in a sensitivity analysis, by performing a network meta-regression with the trial precision as a study-level covariate. If there is strong evidence of small study effects or publication bias, we will clearly report it and interpret all results with caution.

Model implementation

We will fit all models in R. We will fit the pairwise meta-analysis models using the **meta** package.²⁷ All frequentist NMAs will be fit using the **netmeta** package.²⁸ We will

perform all Bayesian analyses using the **R2jags** packages.²⁹ For all Bayesian models we will assume a binomial likelihood for the number of events per treatment arm. We will employ uninformative prior distributions, e.g. $N(0,100^2)$ for all location parameters such as the log-odds ratios of relative treatment effects. For the heterogeneity parameter we will employ the empirical distributions described elsewhere.⁸ We will run multiple chains and assess convergence and mixing of the chains using the Brooks-Gelman-Rubin diagnostic criterion.

Assessing the confidence of evidence of NMA

The quality of evidence obtained by the synthesis of the evidence for each outcome will be separately evaluated using the framework described in Salanti et al.³⁰ and implemented using the CINeMA³¹ (Confidence in Network Meta-Analysis) web application. This will allow to grade the confidence in the results into high, moderate, low or very low.

Discussion

The adverse effects of the antidepressants and their perceived marginal efficacy are major factors contributing to the unsatisfactory treatment duration of antidepressants. These factors are exacerbated by our current inability to predict which drug will cause the fewest adverse effects for a specific patient, and which will work most effectively. This work will provide an in-depth analysis and an insight into the specific adverse events of individual antidepressants. This NMA is a key step in retrieving and understanding all of the information needed to guide the shared decision-making process between patients, carers, and clinicians. It has been widely reported and recognised in the scientific literature.³² Matching patients to individual antidepressants, this will enable clinicians to precisely customise treatment to patients' needs and thus improve their outcome.

Conflict of Interests

SK reports grants from the Mental Health Okamoto Memorial Foundation, Pfizer Health Research Foundation, and KDDI Foundation outside the submitted work.

HI reports lecture fees from Mitsubishi-Tanabe and Yoshitomi.

TAF reports personal fees from Meiji, Mitsubishi-Tanabe, MSD and Pfizer and a grant from Mitsubishi-Tanabe, outside the submitted work; TAF has a patent 2018-177688 pending

All other authors report no conflict of interest

Funding

This research was funded by the National Institute for Health Research (NIHR) Oxford Health Biomedical Research Centre (grant BRC-1215-20005). The funders had no role in the design and conduct of the study; or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgements

AC is supported by an NIHR Research Professorship (grant RP-2017-08-ST2-006). The views expressed are those of the authors and not necessarily those of the UK National Health Service, the National Institute for Health Research, or the UK Department of Health.

.

References

- 1 GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392:1859-1922.
- 2 Cipriani A, Furukawa TA, Salanti G, *et al*. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 2018;391:1357-1366.
- 3 Furukawa TA, Salanti G, Atkinson LZ, *et al*. Comparative efficacy and acceptability of first-generation and second-generation antidepressants in the acute treatment of major depression: protocol for a network meta-analysis. *BMJ open* 2016;6:e010919.
- 4 Cipriani A, Salanti G, Furukawa TA, *et al*. Antidepressants might work for people with major depression: where do we go from here? *Lancet Psychiatry* 2018;5:461-463.
- 5 Cipriani A, Furukawa TA, Salanti G, *et al*. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009;373:746-758.
- 6 Mantel N., Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 1959;22:719-748.
- 7 Efthimiou, O. Practical guide to the meta-analysis of rare events. *Evid Based Ment Health* 2018;21:72-76.
- 8 Turner RM, Davey J, Clarke MJ, *et al*. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International journal of epidemiology* 2012;41:818-827.
- 9 Efthimiou O, Debray TP, van Valkenhoef G, *et al*. GetReal in network meta-analysis: a review of the methodology. *Research synthesis methods* 2016;7:236-263.
- 10 Mavridis D, Giannatsi M, Cipriani A, *et al*. A primer on network meta-analysis with emphasis on mental health. *Evid Based Ment Health* 2015;18:40-6.
- 11 Kendall T, Morriss R, Mayo-Wilson E, *et al*. Assessment and management of bipolar disorder: summary of updated NICE guidance. *BMJ* 2014;349:g5673.
- 12 Wijkstra J, Lijmer J, Burger H *et al*. Pharmacological treatment for psychotic depression. *Cochrane Database Syst Rev* 2015;7:CD004044.
- 13 Barbui C, Cipriani A, Patel V, *et al*. Efficacy of antidepressants and benzodiazepines in minor depression: systematic review and meta-analysis. *Br J Psychiatry* 2011;**198**:11-16.
- 14 Shinohara K, Tanaka S, Imai H, *et al*. Development and validation of a prediction model for the probability of responding to placebo in antidepressant trials: a pooled analysis of individual patient data. *Evid Based Ment Health* 2019;22:10-16.
- 15 Gibbons RD, Hur K, Brown C, *et al*. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Arch Gen Psychiatry* 2012;**69**:572-579.

- 16 Khan A, Kolts RL, Thase ME, *et al.* Research design features and patient characteristics associated with the outcome of antidepressant clinical trials. *Am J Psychiatry* 2004;**161**:2045-2049.
- 17 Rucker G, Schwarzer G. Reduce dimension or reduce weights? Comparing two approaches to multi-arm studies in network meta-analysis. *Stat Med* 2014;**33**:4353-4369.
- 18 Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 2011;**64**:163-171.
- 19 Efthimiou O. A Mantel-Haenszel model for network meta-analysis of rare events (Under review in *Statistics in Medicine*).
- 20 Stijnen T, Hamza TH, Ozdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med* 2010;**29**:3046-3067.
- 21 Efthimiou O, White I. The dark side of the force: multiple testing issues in network meta-analysis and how to address them (under review in *Research Synthesis Methods*).
- 22 Higgins JP, Jackson D, Barrett JK, *et al.* Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods* 2012;**3**:98-110.
- 23 Lancee M, Lemmens CMC, Kahn RS, *et al.* Outcome reporting bias in randomized-controlled trials investigating antipsychotic drugs. *Transl Psychiatry* 2017;**7**:e1232.
- 24 Schroll JB, Penninga EI, Gotzsche PC. Assessment of Adverse Events in Protocols, Clinical Study Reports, and Published Papers of Trials of Orlistat: A Document Analysis. *PLoS Med* 2016;**13**:e1002101.
- 25 Peters JL, Sutton AJ, Jones DR, *et al.* Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *J Clin Epidemiol* 2008;**61**:991-996.
- 26 Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med* 2006;**25**:3443-3457.
- 27 Schwarzer G. meta: An R package for meta-analysis. *R News* 2007;**7**:40-45.
- 28 Rücker G, Krahn U, König J, *et al.* netmeta: Network Meta-Analysis Using Frequentist Methods. <https://CRAN.R-project.org/package=netmeta> (2019).
- 29 Yajima, Y.-S. S. a. M. R2jags: Using R to Run 'JAGS'. R package version 0.5-7. <https://CRAN.R-project.org/package=R2jags>. (2015).
- 30 Salanti G, Del Giovane C, Chaimani A, *et al.* Evaluating the quality of evidence from a network meta-analysis. *PLoS one* 2014;**9**:e99682.
- 31 CINeMA: Confidence in Network Meta-Analysis [Software]. Institute of Social and Preventive Medicine, University of Bern, 2017. Available from cinema.ispm.ch.
- 32 Leucht S, Chaimani A, Cipriani A, *et al.* Network meta-analyses should be the highest level of evidence in treatment guidelines. *Eur Arch Psychiatry Clin Neurosci* 2016;**266**:477-80